DOCUMENT RESUME

ED 145 833 IR 005 264

AUTHOR Hjerppe, Roland
TITLE A Meta-Model for Evaluating Information Retrieval Serviceability.
INSTITUTION Royal Inst. of Tech., Stockholm (Sweden). Library.
REPORT NO TRITA-LIB-6002
PUB DATE Nov 74
NOTE 35p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Design; *Evaluation; *Information Retrieval; *Information Systems; Models; Systems Analysis; *Systems Approach

ABSTRACT

This document first outlines considerations relative to a systems approach to evaluation, and then argues for such an approach to the evaluation of information retrieval systems (ISR). The criterion of such evaluations should be the utility of the information retrieved to the user, and the ISR ought to be regarded as one of three interrelated systems which include also the users and the information producers. The method suggested for describing an ISR system as a service system involves following one query through the system from first to last contact. Parameters identified for each point through the system can be described as either global or local and belong to one of five categories: contextual, serviceability, interface, information service, customer. Questions of cost-benefit and cost-effectiveness are not addressed. (WBC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED145833

REPORT TRITA-LIB-6002 NOVEMBER 1974

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A META-MODEL FOR EVALUATING INFORMATION RETRIEVAL SERVICEABILITY

Roland Hjerppe

Royal Institute of Technology Library
S-100 44 STOCKHOLM 70

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephan Schwarz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

{Published in K. Samuelson et al.: *Evaluation of Information Retrieval Systems and Functions*, NORDFORSK, Stockholm 1974}

IR005264

A META-MODEL FOR EVALUATING INFORMATION RETRIEVAL SERVICEABILITY

INTRODUCTION

The aim of this section is, as indicated by the chapter title, neither to provide a model nor to be a procedural guideline with explicit rules for an evaluation. It is rather an attempt at providing an outline of the process of modeling and designing an evaluation of the serviceability of ISR services. As such it is an indication of how much there is needed a coupling of all the results from "information needs and uses" studies to the design, evaluation and redesign of information services.

A SYSTEMS APPROACH TO EVALUATION

The evaluation of an ISR service is in no way different from the evaluation of systems in general. One should thus proceed by a systematic, unbiased, thorough analysis and description of the system, its related systems and parts, as well as of the evaluation process. A system is always part of other systems or supersystems, and has parts plus subsystems. A system consists of parts (subsystems), relations between these parts and the structure defined by these relations and parts.

On Description

The basic information requirements which must be met in describing a system (Chestnut, 1970), are the following:

Distinguishing qualities of the parts:

A qualitative description of those particular variables and parameters that will be used to characterize the parts and differentiate them from other parts of similar structure and nature.

Magnitude, probability and time span:

A quantitative description of each of the variables and parameters in terms of the ranges and limits of their amplitudes (magnitudes), time span and/or frequency and their degree of certainty (probability) for the situations or conditions of significance to the system during its existence.

Types of relations:

The interrelationships of the parts in various respects such as space, time, causality, logical properties etc.

Structure:

The various structures defined by the parts and their relations.

System dynamics:

What a system does in relation to its supersystem(s).

There is no selfevident, selfasserting and self-existent system, it is rather in the eye of the beholder. Every system is part of many supersystems and each of its subsystems are part of many systems, so in order to be able to isolate one of these one has to define the context and objectives for the description. The description of a

system does thus depend on:

1) who the describer is, his ability for perceiving, characterizing and differentiating the system,
2) why he describes, the reason for and purpose of the discription,
3) where he is, the context within which and the position from which the description is made,
4) how he describes, the language, methodology and termirology used,
5) for whom the description is made, his context, position, knowledge etc, and of course
6) what it is, the system itself.

E.g. a library will be described differently by a librarian, an architect, a user. A librarian will describe the library when he wants to raise funds, differently from when he wants to get new users. A reference librarian will not give the same description as an accuisition librarian. The holdings of the library can be described for a new user by pointing at a catalogue or holdings list or by taking him around and giving him examples. The description of a catalogue when described for a user will differ from when described for a visiting librarian. The description of the library does of course also depend upon the library itself, big or small, general or specialized, public or private.

On Evaluation

Evaluation is a process of finding out how a system performes some of its roles within one of its supersystems. The evaluation process can be devided into the following steps:

I Statement of purpose, extent and context of evaluation, including identification of objectives and allocation of resources for the evaluation process.

II Definition, delimitation and description of the system to be evaluated, including system environment and context.

III Description of the functions and roles of the system, i.e. the objectives it should accomplish.

IV Transformation of these objectives into criteria for evaluation.

V Identification of variables and parameters, the measurement of which can give an indication of satisfaction of criteria.

VI Measurement of variables and parameters.

VII Comparison of results and criteria.

VIII Interpretation and drawing of conclusions.

IX Decision and action (preventive or corrective).

I In defining the purpose of the evaluation it should be stated whether the evaluation is aimed at:

Performance of one system;

in this case difference should be made between evaluation for:

- Design or implementation of new system
- Testing of implemented system
- Monitoring of existing system
- Modification of existing system
- Survival of existing system

Comparison between different systems
(This is much more difficult and will not be discussed here).

The extent of an evaluation can broadly be classified into one of these cases:

- Macro evaluation
- Micro evaluation
- Intermediate evaluation

Macro evaluation is mostly concerned with the question "How well does the system perform?" with regard to certain objectives. In micro evaluation the questions are "How does the system perform?" and "How well do the subsystems perform?" Intermediate evaluation is a mixture of macro and micro evaluation (Westat Research, Inc. 1968).

The context of evaluation can be differentiated into

- Evaluation from above
- Evaluation from within, management or subsystem operational performance
- Evaluation from outside.

In the evaluation of a system from above, the system to be evaluated is defined as a subsystem in a supersystem that includes those other systems that the system is related to or interacts with. In an information service system, an evaluation from above would include the customers and information producers in the definition of the system to be evaluated.

In an evaluation from within the evaluation is usually concerned with the operations of the system's

parts, either from a management, from a coordinating or from a technical operational point of view. Everything outside the system is considered only in terms of its direct influence on and relation to the system.

An evaluation from outside is made from within another system (user) and the system is evaluated with respect to its performance as a subsystem, with the outside system objectives being the basis for criteria. See fig. 5.

II System description has been dealt with on p. 30-32.

III The functions of a system can be stated in terms of objectives e.g. "provide SDI-service to a certain community of researchers". But this is not explicit enough, usually there are also goals formulated such as "Give a weekly list of new references selected from databases A, B, C, according to personal profiles, with an average precision of 0.60 and recall of 0.40" etc. or the goal may be to maximize and minimize certain parameters with stated constraints.

All systems have a (sometimes neither explicitly nor implicitly stated) goal structure. These goals are usually set at different levels, priority and range, and are perhaps not compatible with each other. The goal structure depends on the supersystem of which the system is a part. A subsystem is designed and constructed in order to perform a certain role or fulfill functions in a system. The evaluation of a system is thus a question of establishing how well a system performs toward its

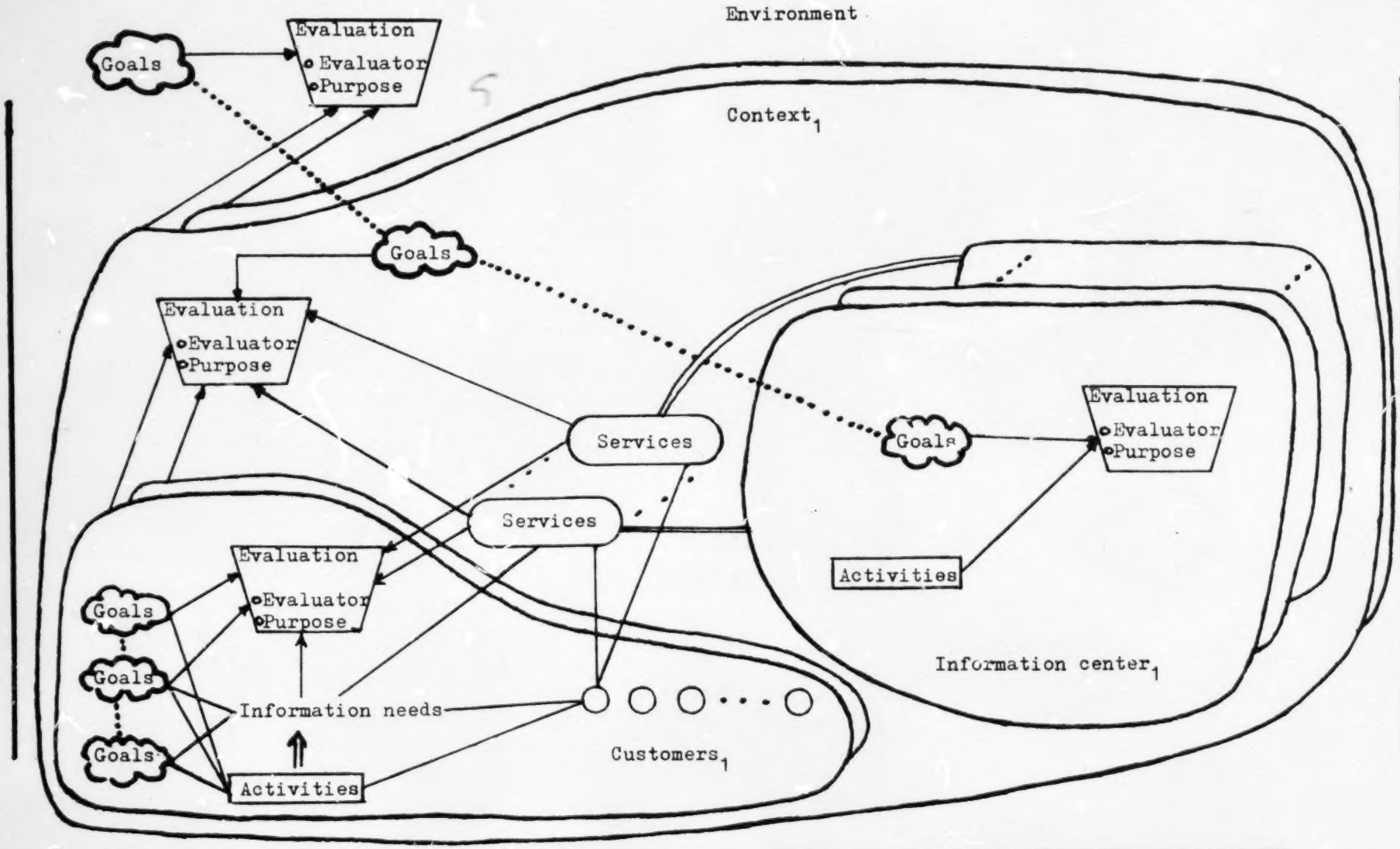
Environment
Goals
Evaluation
Evaluator
Purpose
Context1
Goals
Evaluation
Evaluator
Purpose
Services
Goals
Evaluation
Evaluator
Purpose
Services
Activities
Evaluation
Evaluator
Purpose
Goals
Goals
Goals
Information needs
Activities
Customers1
Information center1

Figure 5

goal and satisfies the goals derived from the supersystem. The accurate definition of objectives and goals is one of the critical steps in evaluation.

IV-V The transformation of these goals and objectives, once verbalized, into criteria is not easy. There is in practice an interdependence between the criteria we choose and the parameters and variables available for describing performance. If we want to transform the objective "easy to use" into certain goals and these into criteria, then we actually have to start with finding out what we mean by "easy to use" and once this is done, how we can measure it. Not until then can we give reasonable criteria.

VI The method to be used for measuring the defined variables is not at all evident from the definition of the variables. A parameter or variable is only something to be measured, how to do it is something else. An example will illustrate this. The volume of an irregular body can be measured, by measuring its linear dimensions, if this is not feasible, another method must be devised, e.g. the volume of water it displaces (cf. Archimedes' principle).

The metrics to be used in measuring and giving the results is also a matter to be considered.

VII-VIII The comparison of measurement results and criteria is quite straight forward but the interpretation of these results and the conclusions that can be drawn will vary. One main difficulty is the problem of multi-dimensional

ranking, i.e. how do you rank systems that need more than one variable for description, e.g. which is better, a retrieval system with 0.75 precision and 0.40 recall or an ISR-system with 0.55 precision and 0.60 recall?

IX The decisions and actions based upon the interpretations will in their turn depend on the objectives of the supersystem and will thus be quite outside the system.

The system and its supersystem, its environment and context, evaluator and evaluation criteria are all closely interconnected and -dependent. Neither of these is self-definable, but together they can delimit the boundaries and aspects of each other and thus be used for definition and description of the parameters which are applicable and adequate for an evaluation of some aspects of a system, see Fig. 6.

SERVICEABILITY AND ITS MEANING

Applying the approach that has been outlined above we find that the traditional way of evaluating ISR-systems using measures of precision and recall might be relevant for those responsible for the maintenance of indexing systems and that these measures which might be useful and applicable when considering small, well-defined and slowly growing collections are quite inadequate in the dynamic, ever-changing, almost chaotic information generation, flow and decay of to-day's reality. An analogy can illustrate this:

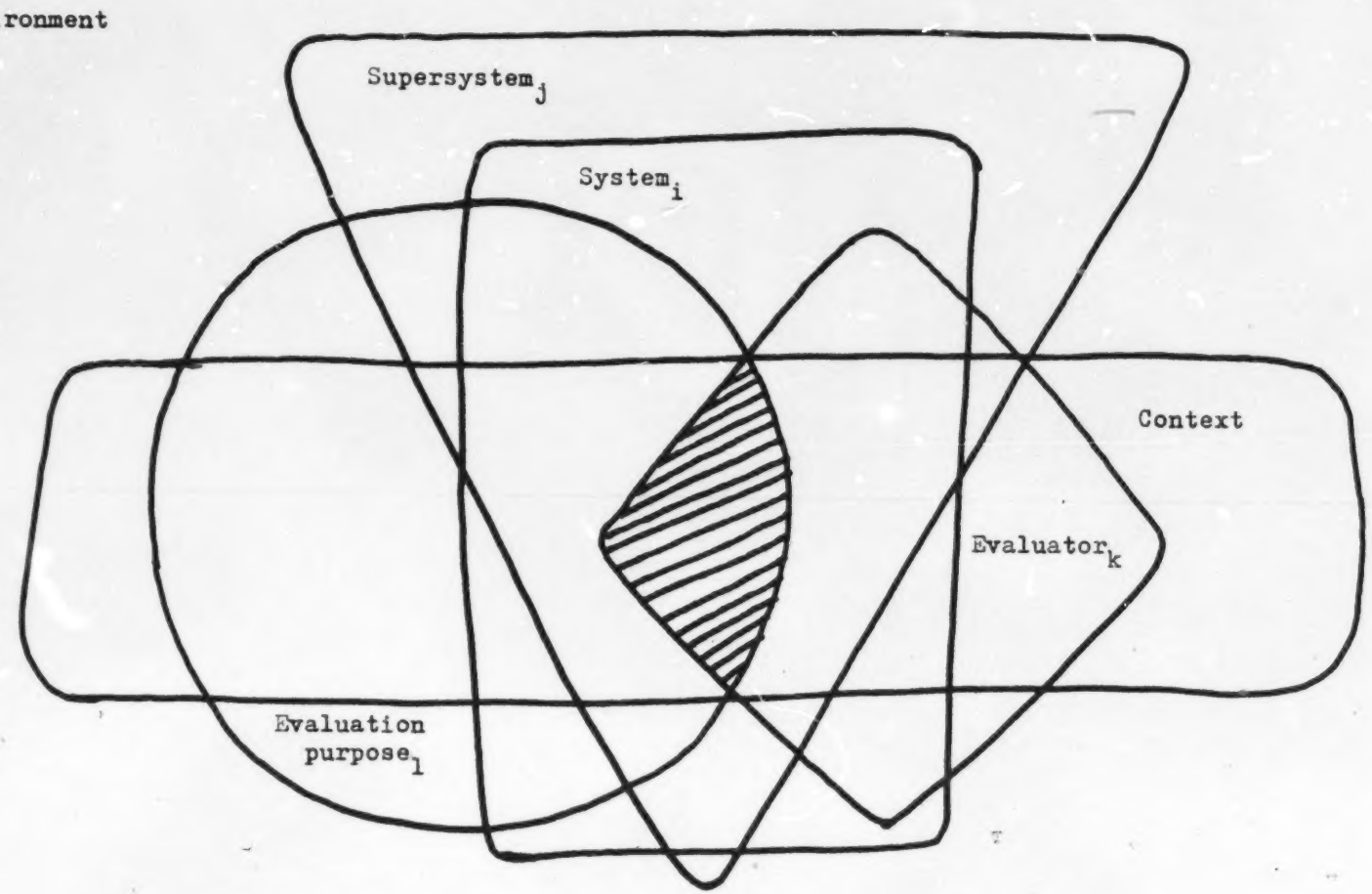
Environment
Supersystem$_j$
System$_i$
Context
Evaluator$_k$
Evaluation purpose$_l$

Figure 6

There are many methods and techniques that can be used for fishing. There are many different species of fish and within species there are great individual variations. There are many different places where you can fish, there are different fishers with different purposes and goals. The evaluation of the fishing process using measures such as percentage of goldfish of total catch and percentage of goldfish cought out of the total population of goldfish, is relevant only if you are trying to catch goldfish in a small pond, with semistatic population, i.e. either the fishes are neither dying nor procreating, or you mark and classify and count each fish before allowing it to enter the pond. But when you go fishing in a stream or roam the high seas on a trawler, then you will need different measures for the evaluation of your activities

Most evaluation studies have so far been concerned with evaluation from within. There exists on the other hand many studies on "information needs and uses" as witness all the volumes of Annual Review of Information Science and Technology, but all the knowledge in these studies has not as yet been used as one of the possible bases for a systematic evaluation from outside.

ISR-systems can be regarded as service functions and evaluated as such. In that case the evaluation should be made from the outside considering external properties from the user's point of view and from above. The inner operations are of course extremely interesting to the people inside the system, but a customer mostly considers an ISR-system to be a black box providing him with information and he is only interested in those characteristics of that black box which he perceives directly and which

interfere, positively or negatively, with his own activities. For a user the evaluation should in fact be made with orientation towards his own utility of the information retrieved. This means that the evaluation should be result-oriented, and that the result is not the printed list of references but the decisions made and actions taken on the basis of information contained in the documents referred to.

In evaluating an ISR-system from above we are concerned not only with evaluating the operational characteristics of the ISR-system but also with the effect this will have upon the user population. The question in this case is not "how well does the system perform" but "how well do the interrelated systems perform jointly in their attempts at fulfilling the goals and objectives of the supersystem". In other words, an ISR-system and a research system (users) and information producers are considered as one system, the goals of which will be of higher level and priority. Independent optimization of each part, subsystem, might well be inconsistent with the optimization of the total system. When considering an ISR-system from above we find that the information delivered, although important, is not as important as the act of providing the information.

The evaluation of a service providing system is not as much concerned with <u>what</u> the system provides as with <u>how</u> it does it and to <u>whom</u>. Which the pieces of information are is not relevant except to the customer (user). Naturally the selection of the right pieces is important, but this is just one step in the service-providing process. An ISR-system, with high figures of recall & precision, that is cumbersome to use, or slow in answering, or geographically out of the way will not be used, and the amount of usage

gives an indication of user acceptance and thus also of the serviceability of the system. What good is a perfect selection process that no one uses? On the other hand only service and no content will attract only new customers and this only for a while. There is thus needed a balance between service and content.

The remaining part of the chapter text will be devoted to giving very briefly one possible characterization of an ISR-system for the purpose of evaluating it from above, i.e. the view of some funding agency or coordinating management. The characterization is aimed at evaluating the performance of one system and should be applicable to all subdivisions. The aim is at providing outlines for intermediate evaluation so the degree of resolution and detail is not the same throughout. As this is only a sketch it will not be concerned with the formulation of criteria nor with measurement nor comparison and interpretation. Additional research is desirable before the description and its variables and parameters are used for an actual evaluation.

EVALUATION OF SERVICEABILITY

There are many possible angles of attack in attempting to describe an ISR-system as a service-providing system. The one chosen here presupposes an existing system in a steady or semi-steady state, i.e. the system has existed for some time, and is not at the present time undergoing any major changes, so all transients are negligible. The system considered has three parts, a user community, an information-providing service, and information producers.

Only serviceability with regard to the user community will be of interest. The service to the information producers is of interest only when interfering, positively or negatively, with the service to the user. There might of course be incompatibilities between the goal for the information system in wanting to provide as many, varied customers as possible with service and the specific customers' need for personal services.

The angle that has been chosen is to follow one query, one service-demand, through the system, from first to last contact, from the users becoming aware of the information service to the termination of contact with respect to this particular service demand. The service-demand and -delivery process can be divided into the following sequential steps:

1. Information about the information service
2. First contact
3. Query formulation
4. Query processing
5. Output
6. Feedback
7. Acquisition
8. Termination

Naturally, one specific customer need not pass through point 1 every time, and one specific query might cycle through points 4-7 for a long time.

With every point and every transition between points there are associated sets of variables and parameters. In Fig. 7 is indicated how the information need of the user will influence this process and how this process

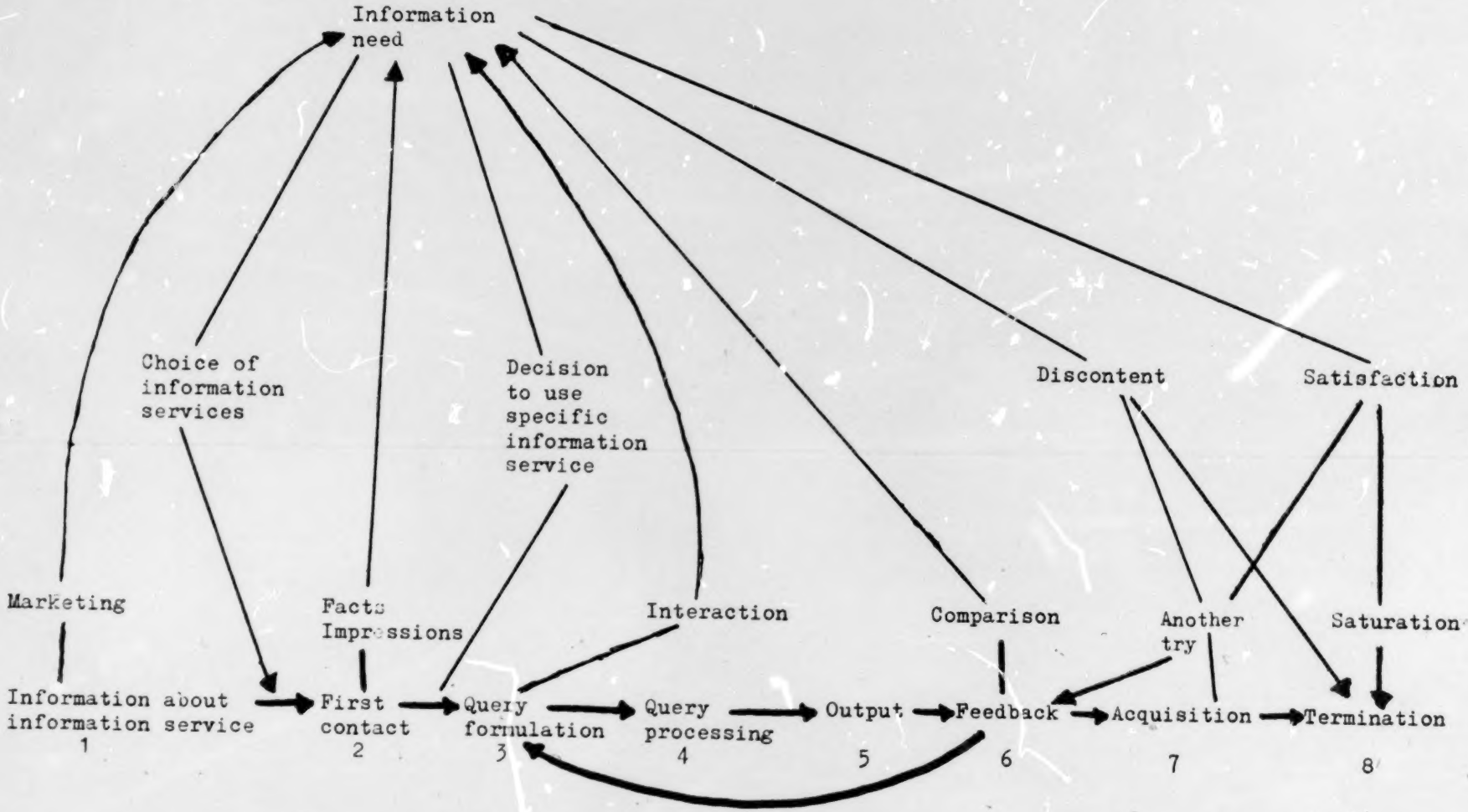
Information need
Choice of information services
Decision to use specific information service
Discontent
Satisfaction
Marketing
Facts Impressions
Interaction
Comparison
Another try
Saturation
Information about information service
1
First contact
2
Query formulation
3
Query processing
4
Output
5
Feedback
6
Acquisition
7
Termination
8

Figure 7

will initiate actions from the user.

The information need of the user is only one of the many parameters required when we attempt to describe the complex processes of interaction between user and information service that occur at every point and every transition in the sequence.

The parameters needed for a description of all these processes will be of two types:

- Global parameters that are applicable at any point or any transition
- Local parameters that apply only in a certain point or transition.

The parameters can also be classified as belonging to one of the following categories:

- Contextual parameters
- Serviceability parameters
- Interface parameters
- Information service parameters
- Customer parameters

The relationships between these sets of parameters and between these and the service demand sequence is indicated in Fig. 8.

The contextual parameters will in the specific evaluating cases determine the selection of parameters out of the total set possible for each of the other categories of parameters.

Obviously, all the parameters describing the customer are global and the values of these parameters for a

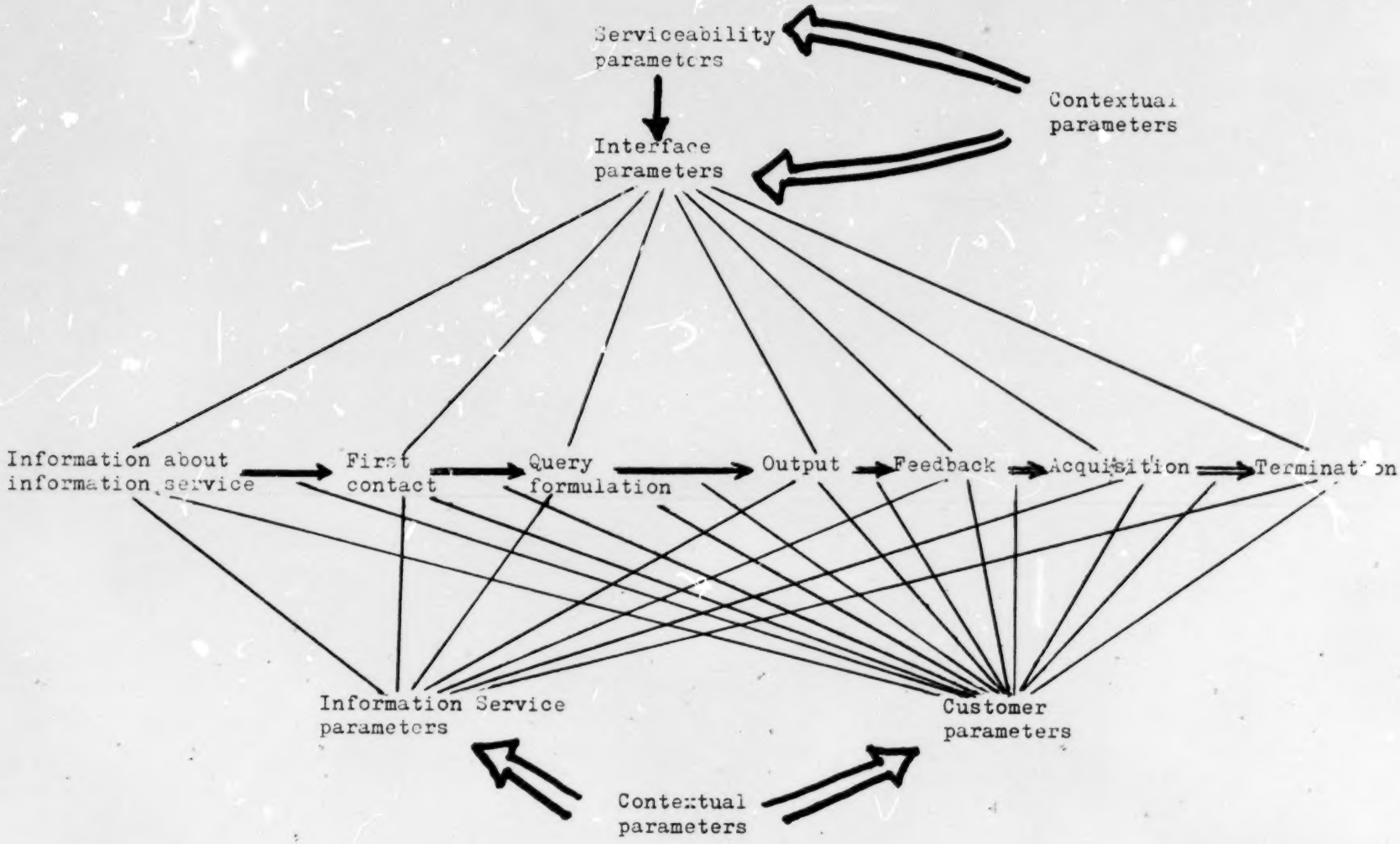
Serviceability parameters
Contextual parameters
Interface parameters
Information about information service
First contact
Query formulation
Output
Feedback
Acquisition
Termination
Information Service parameters
Customer parameters
Contextual parameters

Figure 8

specific customer will in a specific evaluation situation influence the selection of other local parameters and the ranges of values of these applicable for an evaluation of the service from that customer's point of view.

The information service parameters apply in the specific points of the service demand sequence and describe the capabilities and potentials of the information service.

The interface parameters specify the interaction between the customer and the information service and are thus the parameters that in each point characterize the service delivery.

The serviceability parameters are used in every point of the service demand process and in every point applied to every interface parameter in order to evaluate the service delivery process.

The model can be summarized very briefly in the following way:

In a specific case of evaluating the serviceability of a certain information service with regard to a certain user or group of users there are contextual parameters, the values of which will have determined the selection of parameters, and in some cases their ranges of values, that are pertinent for describing the users, the information service, the serviceability and the interface. The serviceability is evaluated by different measures in every point of the service demand sequence. In every point of the service demand sequence the interface parameters describe the interaction between the information service and the customer. This interaction depends on the

characteristics of the customer and the information service with regard to the point in question. The serviceability parameters are then used to evaluate the service delivery in that point for each of the interface parameters.

LISTS OF PARAMETERS

The following list of parameters represent a fair attempt and is open for additions. Many of the parameters are not easily quantifiable and for many of those to which a scale might be assigned the problem of determing the values within that scale with accuracy and reproducibility is not simple.

Contextual Parameters

The parameters that describe the context will determine the selection of and ranges of variation of all other parameters.

ORGANIZATION	Internal structure External affiliations Type, public - private
GEOGRAPHICAL LOCATION	
ECONOMY	Budget, fixed or flexible Funding, profit or cost-recovery or subsidized
OBJECTIVES	Goals
COMPETITION	Roles

Customer Parameters

From a serviceability point of view the customers can be described with the following parameters.

MOTIVATION	of requester
	Use to be made of information
INFORMATION NEED	Speed requirements
	Finality requirements
	Vitality requirements
	Depth (specialization)
	Width (generalization)
	Type, ideas, problem solution, data
	Actuality, age of document
	Time span
	Retrospective search - SDI
	Subject area
TYPE	Occupation
	Education
	Affiliation, industry-university
	Position
RESOURCES	Time
	Money
PRIOR KNOWLEDGE	Awareness of state-of-the-art
	Acquaintance with information
	Services

Information Service Parameters

SIZE	Resources Personnel Number of users Volume of information handled
TYPE OF SERVICE	Retrospective searches - SDI - Reviews Back-up, referral
COVERAGE	Depth Width Time span Subject areas
TIMELINESS	Service Information
RELIABILITY	Processing (MTBF, MTTR)
CREDIBILITY	Does the system do what it says it can do?
REPUTATION	
CAPACITY	Step response
BROWSING CAPABILITY	
COMPATIBILITY	Internal External

REDUNDANCY	Overlap, coverage and services, internal and external
INTERMEDIARIES	Need Availability
COST	
PERFORMANCE	Precision Recall

Interface Parameters

The parameters characterizing the interface in the service demand and delivery process are the following, taken point by point.

1 Information about the information service

MEDIUM	Paper, film, speech
TYPE	Advertisement - journal article
CHANNEL WIDTH	Personal communication - broadcast
APPEARANCE	Layout, illustrations, colours
ACTUALITY	of information
COMPREHENSIVENESS	Popular - detailed
COMPLETENESS	Alternatives Limitations Omissions

2 First contact

MEDIUM	Letter, telex, telephone, personal visit
RESPONSE	Speed Effort Empathy, personal - impersonal Direct contact - intermediary
PERSUASIVENESS	Special offers, free trial Marketing
CONVENIENCE	
UNDERSTANDING	of service demand user need user situation

3 Query formulation

FORMALITY	Free form - special language, symbols Narrative form - references
ASSISTANCE	Amount needed and available
AIDS	Thesauri, frequency lists Printed versions Manuals
SEARCHABLE ELEMENTS	Title, author, keyword, ...
OPERATORS	Logical, weighting, nesting, ...

RANKING	Ordering schemes or algorithms
ADAPTATION	one profile for all databases vs. a special profile for each database
EXTENDABILITY	Profile hierarchies, macro profiles

5 Output

MEDIUM	
BIBLIOGRAPHIC LEVEL	Reference, abstract, ...
LAYOUT	
FORMAT	
LEGIBILITY	
REGULARITY	
EXPLANATIONS	
CUMULATION	Selected cumulated printout possibility

6 Feedback (in addition to those under 3 Query formulation)

TYPE	Active - passive
REGULARITY	Continuous - discrete

AIDS — Statistics
Updated profile printouts
Search characteristic curves

MEDIUM

7 Acquisition of interesting documents (back-up service)

ROLE — In-house - external

FORMALIZATION — In ordering, tabs - order forms

SPEED — of delivery

ACCURACY

MEDIUM — of acquired documents
Copies, tear sheets, loan, microforms

COST

8 Termination

RESPONSE

RETENTION OF FILES

RESTART

Serviceability Parameters

AMOUNT OF CHOICE	Type, range of alternatives, options
USER EFFORT	
CONVENIENCE	to user
INTERACTION LEVEL	
EDUCATION POSSIBILITIES	for user
ACCESSIBILITY	for user
ADAPTABILITY	to user
RESPONSE TIME	
COMMUNICATION MODE	
DOCUMENT BACK-UP	
HOSPITALITY	to different service demands
DIFFERENTIATION	between closely equal service demands

CONCLUSIONS about SERVICEABILITY

The above proposed list of parameters is neither complete nor unconditional. Most of the parameters have been identified and in some instances illustrated. The proper definition of the parameters, the measurement methods and associated metrics must depend on a consensus on the appropriate sets of parameters. Naturally not all parameters are equally important, but even a minimum set of parameters needs to be ranked somehow. The problem of multidimensional ranking, although discussed in the literature, has not received enough attention from evaluators and designers.

It does seem that for a very simple evaluation of serviceability, either the amount of choice or user effort or preferrably both are the serviceability parameters that can best be applied in any point of the service sequence and for all interface parameters.

The costs and economics have only been mentioned in brief. This does not imply that these are in any way negligible, rather the contrary. Every activity, and non-activity has an associated cost in terms of resources such as personnel, time and money. A wide range of alternatives in every interface parameter can be equated with having a big reserve supply in stock and does thus involve large costs compared with a small supply of a small range of items.

The balance between service and costs is not easy but is to some extent determined by contextual parameters for information services. Within a fixed budget then the problem arises which service parameters to rank as the

highest. All these questions of cost-effectiveness, cost-benefits, benefit-effectiveness, service-cost, service-benefits etc can not be discussed here but their importance remains. Serviceability as such is not, however, directly concerned with costs.

REFERENCES

H. Chestnut: Information Requirements for Systems Understanding {*IEEE Transactions on Systems Science and Cybernetics 6*(Jan. 1970)no. 1}

Westat Research Inc.: *Procedural Guideline for the Evaluation of Document Retrieval Systems* {Office of Scientific Information Service, National Science Foundation, Washington, D.C., (Dec. 1968) PB 182 711}